DOCUMENT RESUME

ED 320 939 TM 015 188

AUTHOR Smith, Michael W.
TITLE Using Rasch Analysis To Design and Interpret a Test of Students' Understanding of Irony in Poetry.
PUB DATE 90
NOTE 41p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Testing; High Schools; *High School Students; *Irony; Item Response Theory; Literary Devices; *Poetry; Pretests Posttests; Rating Scales; Reading Tests; *Scoring; *Test Construction; Test Interpretation
IDENTIFIERS *Rasch Model

ABSTRACT

The insights provided by Rasch analysis of results from a literature test were explored. Students in grades 9, 10, and 11 (n=261) responded to a 28-item test before they received one of three treatments: (1) direct instruction, based on research on metacognition in reading, which attempts to give conscious control of strategies used to understand irony; (2) a tacit method, which seeks to have students develop their own strategies through extended practice with the genre; and (3) no treatment. The test made four statements about each of seven poems, five of which were ironic, and asked respondents to strongly agree, agree, disagree, and strongly disagree with each statement. A revised posttest with additional ironic poems was taken by 253 students. Test analysis suggested that the interaction of confidence and understanding in some students made dichotomous scoring preferable to step scoring. Analysis, especially of the new poems on the posttest, suggested that the belief under attack as well as the difficulty of the language in the poems contributed to students' difficulties in understanding irony. Analysis of pretests and posttests indicated that experimental treatments had equal effects on students regardless of initial ability, although the direct treatment may have been more beneficial for the least able students and for the most difficult tasks. The test proved to be a reliable measure of students' understanding of irony in poetry. Results illustrate the benefits of Rasch analysis in test construction, interpretation, and scoring. Five data tables are provided, and the test is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320939

Using Rasch Analysis to Design and Interpret a Test of Students' Understanding of Irony in Poetry

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
MICHAEL W. SMITH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Michael W. Smith
University of Wisconsin-Madison
Department of Curriculum and Instruction
225 N. Mills Street
Madison, Wisconsin 53713

Running Head: RASCH ANALYSIS OF A TEST OF IRONY

TM015188

BEST COPY AVAILABLE

Abstract

This paper is an analysis of the results of 514 responses to a test of students' understanding of irony in poetry. 261 ninth-, tenth-, and eleventh-graders responded to a 28 item test before they received one of three treatments: (1) a direct method, based on research on metacognition in reading, which attempts to give students conscious control of the interpretive strategies experienced readers use to understand irony, (2) a tacit method, which seeks to have students develop their own strategies through extended practice with the genre, and (3) no treatment. The test made four statements about each of seven poems, five of which are ironic, and asked respondents to strongly agree, agree, disagree, or strongly disagree with each statement. 253 students took a revised posttest that included two additional ironic poems. The analysis of the test suggested that the interaction of confidence and understanding in some students made dichotomous scoring preferable to step scoring. In addition, the analysis of students' responses suggests that ironic items were significantly more difficult than non-ironic items and that non-ironic items in poems containing irony were significantly more difficult than items on poems not containing irony. Subsequent analysis, especially of the measures of the two new poems on the posttest, suggests that the belief under attack as well as the difficulty of the language of the poems contributed to students' difficulty in understanding the irony. Finally, analysis of students whose measure changed by one or more standard errors suggested that both experimental treatments had similar effects on students regardless of their initial ability. Also, the analysis suggested that the direct treatment may have a greater benefit than the tacit treatment for the least able students and on the most difficult tasks. However, the direct treatment did appear to have some risk, for more students' performance significantly decreased with this treatment. Further, the analysis suggested that both experimental treatments would be improved by additional instruction in recognizing when a text is not ironic.

Designing a test that evaluates the quality of students' responses to literature puts one on risky theoretical ground, for it is not a popular position to say that some responses to literature are better than others. However, the insights Rasch analysis of such a test can give more than justify the risk. My paper will begin by presenting a justification for measuring the quality of students' responses to literature. It will continue to present key findings in three areas: test construction, understanding the variable, and understanding the quality of instruction.

A Rationale for Measuring the Quality of Students' Responses

Why is it important to evaluate responses to literature? Cooper (1985) explains that evaluation may be important to prove the worth of our programs to those on the outside--principals, superintendents, and the public. Marshall's (1989) study of six experienced English teachers who had reputations within their departments for being excellent teachers suggests that we may have to evaluate responses in order to be persuasive to those within the profession. As part of his study, Marshall asked the teachers to explain the purposes that drive classroom discussions of literature. He notes what he calls a "doubleness" in all of the teachers' responses to this question. For example, one teacher explains that a good discussion "is most like a jam session. I'm not sure that is the right phrase, but you know what I mean. A good classroom discussion is most like an interaction of interested

minds on a common topic" (p. 12). Later in the interview she explains that the purpose of a good discussion is

> to get the class from their initial, personal responses and questions to the beginnings of an analysis or the beginnings of looking at ways that an analysis can take place. I'm using. . .the discussion to lead them to find ways to get to a closer analysis of the text. (p. 12)

Marshall summarizes the teachers' beliefs about their roles and about the purposes of discussion as follows:

> On the one hand, teachers are to be merely facilitators; on the other, they are to make certain that the discussion "goes somewhere." One the one hand, they are to provide many students with the opportunity to speak; on the other,they are to make certain that those students "see" certain things about the literature they have read. On the one hand,discussions are "interactions"; on the other, they are occasions in which teachers take students further and deeper into an analysis of the text. (p. 16)

Marshall's analysis of the patterns of teacher and student talk in the classrooms of these teachers clearly reveals that the goal of "getting somewhere" dominated the goal "interaction." It seems, then, that to convince these teachers to consider new methods, we must be able to establish that these methods will help students "get somewhere." And that means evaluation.

What's a researcher to do? On one hand stand the people who have the opportunity to do something with the results of our research: the principals, the department chairs, the teachers themselves. And many of these people need evidence of "improvement" to be persuaded. On the other hand prevailing critical theories challenge the legitimacy of judging the quality of responses. As Booth (1975) explains, "One hears it said these days that understanding is not possible in any normative sense; [every person] constructs [his or her] own meanings and the more variety we have the richer we are" (p. ix). The work of Holland (1975) Bleich (1975), Fish (1980) and many others would support such a view.

That is why I decided to consider irony, for irony is one domain in which a strong justification for judging the quality of responses can clearly be made. Booth explains that "we should marvel, in a time when everyone talks so much about the breakdown of values and the widening of communication gaps, at the astonishing agreements stable ironies can

produce among us" (p. 82). And further:

> I spend a great deal a of my professional life deploring "polar" thinking, reductive dichotomies, either-or disjunctions. And here I find myself saying that only in strict polar decisions can one kind of reading be properly performed. On the one hand, some of the greatest intellectual achievements seem to come when we learn how to say both-and not either-or, when we see that people and works of art are too complex for simple true-false tests. Yet here I am saying that some of our most important literary experiences are designed precisely to demand flat and absolute choices, saying that the sudden plain irreducible "no" of the first step in ironic reconstruction [rejecting the surface meaning of a text] is one of our most precious literary moments. (pp. 128-129)

We make these polar decisions as a matter of course in our everyday lives. When I spill soup on my tie and say, "Nice job, Michael," I expect that my hearers will realize that I am criticizing my clumsiness rather

than praising my artistic ability. As Booth explains, we behave similarly when we read ironic literature. When the speaker of "An Unknown Citizen" asks, "Was he free? Was he happy?" I expect that every experienced reader understands that Auden would answer no. And I would further expect that every teacher would agree that a treatment that helps students come to such an understanding would be desirable. Because I chose a type of literature "designed precisely to demand flat and absolute choices," I was theoretically justified in writing a test to evaluate students' responses to that literature. This paper reports what I learned from developing that test.

Method

Participants

Two hundred and sixty-one students responded to a 28 item test before they received instruction in understanding irony in poetry using one of three treatments: (1) a direct method, based on research on metacognition in reading, which attempts to give students conscious control of the interpretive strategies experienced readers use to understand irony, (2) a tacit method, which seeks to have students develop their own strategies through extended practice with the genre, and (3) no treatment. (See Smith [1989] for an explanation of the treatments.) The test made four statements about each of seven poems, five of which are ironic, and asked respondents to

strongly agree, agree, disagree, or strongly disagree with each statement. 253 students took a revised posttest that included two additional ironic poems.

The Test

The test follows two suggestions of Cooper (1985). He argues that a promising approach for the study of literature is the "literary features test," one which "looks in detail at a student's ability to deal with selected features of a literary text." (p. 318) (See the appendix for a copy of the test.) He also explains the importance of testing students on texts they have not studied before, for "it keeps the testing/evaluation focus on the reader's ability to read insightfully, not on his or her ability to recall what was heard in class or read in literary criticism." (p. 309)

The test is also guided by my understanding of how people interpret irony. Wright and Stone (1979) explain that every test must begin with a sense of the variable the test designer wishes to measure. I designed my test with the belief that the simplest aspect of understanding irony in poetry is to recognize when irony is not present. Consequently, I included two poems that were not ironic in my test. I believe that the next level of understanding is to recognize what information in an ironic poem is not ironic. One must rely on this information in order to reconstruct ironic meanings. Therefore, in three of the five ironic poems I included questions

about information that was not under dispute. I believed that the final level of understanding irony in poetry involves reconstructing ironic meanings in poems whose features, for example, syntax, imagery, and vocabulary, make them increasingly difficult. (Later I will explain how my results caused me to modify this belief.) To test this I included poems that I thought would provide few problems with syntax (e.g., "The Unknown Citizen) and poems whose difficulty would, in my opinion, make their irony more difficult to understand (e.g., "Boom!"). In summary, the initial test makes four statements about each of seven poems, five of which are ironic. Within the ironic poems, some statements require the respondents to reconstruct the meaning in order to answer, while other statements do not require reconstruction.

In addition to being consistent with my understanding of irony, this format is useful for another reason. Were all the statements of the test ironic, students might simply be able to figure out that they should disagree with all the surface meanings. In that case test scores would measure students' ability to figure out the key to a test, not their ability to understand irony. The initial test, then, had twenty-eight items. Respondents were asked to strongly agree, agree, disagree, or strongly disagree with each statement. I intended that each statement would be answered with a strongly agree or strongly disagree response. I piloted a

first draft of the test with seven graduate students in English. I also asked them to write a justification for each of their responses. Most of the variation in response was due to ambiguities in the questions. I then revised the test trying to eliminate these ambiguities. Three experienced English teachers and one graduate student in English took this second draft of the test. One reader gave the desired response to each of the questions. Another had three responses that differed from the desired responses. The third reader had four responses that differed from the desired responses. The fourth reader had six responses that differed from the desired responses. These thirteen variations, however, did not reflect different understandings of the poems; rather the differences occurred when the respondents only agreed with a statement to which the desired response was strongly agree, or only disagreed with a statement the desired response to which was strongly disagree. The similarity of the experienced readers' responses corroborates Booth's statement that stable ironies produce "astonishing agreements" among readers.

After I refined the test, I presented it to the class of 28 honors freshmen whom I used as subjects when I taught the materials for my pilot study, 19 freshmen from the average track, and a class of 16 juniors and seniors. I did a Rasch analysis of the test to evaluate its effectiveness as an instrument and revised questions with fit problems. I piloted this version of

the test with 67 students. I did a Rasch analysis of these data to evaluate the effectiveness of the pretest. Only one of the pilot questions provoked a statistically significant fit problem. I discovered that two of my questions on one poem asked essentially the same thing. I revised one of those questions and two others whose fit statistics approached significance. The final draft of the pretest was administered to the 261 students who began the study.

Results

Evaluating the Test

The test proved to be a reliable measure of students' understanding of irony in poetry. Most importantly, the person fit statistics suggest that the test fulfills the conditions of the model when I scored it regarding students' responses as either right or wrong. That is, if the desired response was strongly agree, I counted both strongly agree and agree as correct and if the desired response was strongly disagree, I counted both strongly disagree and disagree as correct. When I used this scoring system to examine the 514 total responses to the test, only fifteen students had infit statistics of more than two, thus signaling performance that would not be predicted by the probability model.

This was not true in all of the scoring systems that I tried. In my initial analyses of the 263 pretests, I used step scoring. Step scoring

assumes that the response that students chose depended only on their understanding of the poems. That is, it assumes that students made extreme responses only when they totally understood (or totally misunderstood) the poem. However, that assumption proved to be unjustified. Whereas fewer than 3% of the students misfit with the dichotomous scoring, over 10% of the students on the pretest misfit when I used the step scoring system. I hypothesized that the interaction between students' confidence or their lack of confidence and their understanding accounted for these misfits.

To test this hypothesis, I scored the responses in yet another way. I calibrated the responses giving the extreme responses (strongly agree and strongly disagree) a one and giving the moderate responses (agree and disagree) a zero. Only 32 of the 263 students received positive measures in this scoring system, and 24 of these students misfit on the pretest. These analyses suggest that the determining factor in their response was their self-confidence rather than their understanding of irony in poetry. Therefore, I decided to use only the dichotomous scoring in analyzing my results. Perhaps the most important insight into test construction I gained from developing the test is that the possible interaction between confidence and understanding made step scoring unjustified.

The test also made me aware that students' test-taking strategies may affect the fit of an item. The first question to "Sonnet 130" showed a

positive infit of 3.04. The question reads:

> The author believes that the woman described in the poem has unattractive eyes. SA A D SD

I hypothesized that this question misfit because of its position. Students are taught to read test questions before the read the text to which the questions refer. Students who practiced this strategy may have answered the question after reading only the first line of the poem: "My mistress' eyes are nothing like the sun." If they answered the question after reading only the first line, they would not have had a chance to realize the irony of the poem. I theorized that the infit would go down if I put the question last (which I did on the posttest), for in that position students would be less apt to respond without having read the entire piece. In fact, the item had an infit of 2.35 on the posttest, though that decrease could be a function of the instruction or the rehearsal effect of the test as well as its new position.

Understanding Irony

Having finally developed a test that I could use with confidence, I could address important instructional issuues. As Wright and Stone (1979) note, once a test's items have been validated,

> . . . it becomes practical to turn our attention to a [illegible] examination of the calibrated items to see what

> they imply about the possibility of some variable of useful generality. We want to find out whether our calibrated items spread out in a way that shows a coherent and meaningful direction. (p. 83)

If that direction is coherent, it must match, at least to some extent, the test designer's understanding of the variable that informed the construction of the test. My results both supported and refined my conception of what it means to understand irony in poetry.

Table 1 displays the item calibrations. The mean error for these calibrations is .12. I have ordered the items from the easiest to the most difficult. The I in parentheses indicates that the item is ironic. The 1-4 indicates to which of the poem's four items the calibration refers.

Insert Table 1 about here.

Table 2 contains summary statistics of various groups of items.

Insert Table 2 about here.

The data support my belief that, in general, the items that were not ironic were easier than those that were ironic. There was a statistically

significant difference between the groups ($t = 4.25$, $p < .001$).

It is especially instructive to note that "Sonnet 18" is much easier than "Sonnet 130." The mean difficulty of the items on "Sonnet 18" is -.95, while the mean difficulty of the items on "Sonnet 130" is .05. To a great extent, the other aspects of a poem that may increase its difficulty are controlled as Shakespeare authored them both. I attribute the difference in difficulty to the fact that "Sonnet 130" is ironic.

Ironic items appear, in general, to be more difficult than non-ironic items. However, the item calibrations establish that within the ironic items there is substantial variations in difficulty. Initially, I theorized that the difficulty of ironic items would depend on the nature of the poem, its syntax, imagery, etc. However, I underestimated the importance of another major factor, the belief the irony attacks. Booth explains that, "Every reader will have the greatest difficulty in detecting irony that mocks his own beliefs or characteristics." (1974, p. 81)

I believed that Auden's "An Unknown Citizen" would be a relatively easy poem because its syntax and vocabulary are relatively easy. However, the final question, "The author believes that the reader should approve the kind of life the citizen led," was the second most difficult on the pretest.

To check to see if students' prior beliefs could explain this, I asked one class each of ninth, tenth, and eleventh graders who did not participate

in the study to repond to this statement: "Living a comfortable life without controversy is desirable." This statement is one of the ideas that Auden attacks in his poem. Nine students strongly agreed with this statement. Thirty-eight agreed. Twenty-one disagreed, and no one strongly disagreed. I also asked students to respond to this statement: "In a war the officers who plan the strategy are the true heroes, not the soldiers who carry it out." Sassoon attacks this position in "Base Details." Only one student strongly agreed, 4 agreed, 35 disagreed, and 28 strongly disagreed.

Assuming that the students in my study had similar views, a majority went into "An Unknown Citizen" holding a belief that is a subject of Auden's irony. That could explain why the question was more difficult than I anticipated. Again assuming that the students in my study held similar beliefs, they were predisposed to be sympathetic to Sassoon's position, and this predisposition made the irony relatively easy to understand.

Of course, the belief being attacked is not the only factor that contributes to making irony difficult to understand. The final question to "Boom!" was by far the most difficult on the test. This question reads: "The author believes that if people are fortunate it is because God is watching out for them." When I asked students to respond to the statement: "If people are fortunate it is because God is watching out for them," 4 strongly agreed, 30 agreed, 26 disagreed, and 8 strongly disagreed.

Assuming that the students in my study held similar beliefs, slightly more than half of them were predisposed to accept Nemerov's irony. Many fewer were prepared to accept Auden's irony, yet the item on "Boom!" was much more difficult than the item on "An Unknown Citizen." Nemerov's allusions could explain the difficulty of the item on "Boom!" To understand Nemerov's belief, one must understand the significance of Job and Demian and Karnak and Nagasaki. While I gave a note to each of these allusions, the note alone cannot explain all of the associations these allusions have for an experienced reader.

It appears, then, that the belief that is the subject of the irony and the intrinsic difficulty of the poem are the factors that make irony difficult to understand. I tested this hypothesis when I selected the two new poems to be included in the posttest.

The pretest revealed that the test was slightly too easy for the population. The logit measures of 80 students fall above the logit measure of the second most difficult question. This would make it difficult to measure their ability, especially if they improved with instruction. I selected two poems, Ransom's "Dead Boy" and Donne's "Communitie," to increase the number of difficult questions. I selected "Dead Boy" because it has some of the most difficult imagery of all the poems on the test, for example, "the world of outer dark" in line four. Students also have difficulty

with archaic language like that of "Communitie." In addition, each poem had a belief or characteristic that is common among students as a target of its irony. Ransom clearly is not sympathetic with the mourning of the relatives, the "deep dynastic wound" they feel and its expression in tears for the boy. I believe, on the contrary, that most students would be sentimental about the death of any young person and consequently would be sympathetic to the mourning of the relatives. A target of the irony of "Communitie" is the false logic of the speaker, a problem that characterizes much of the writing that students do as well.

I predicted that the poems would be difficult for the students and that within each poem the most difficult item would be the one that ironized a belief or characteristic that I perceive students as having. The data supported my predictions. Using the average item measure as an index, "Communitie" was the most difficult poem, and "Dead Boy" was the third most difficult, behind "Boom!" Further, within each poem the most difficult item was the one I predicted. On "Dead Boy" whereas the students had difficulty in realizing that Ransom did not share the relative's feelings of mourning, they found it much easier to realize that Ransom did not admire these relatives. In the first case, Ransom's belief may run counter to that of many students. In the latter case, students could not have been predisposed to have any feeling for the relatives. On "Communitie" wheras the students

had difficulty recognizing that Donne is poking fun at the false logic of the speaker, they found it much easier to recognize Donne's ironic undercutting of the speaker's attitude toward women. This can be explained when one realizes that students themselves are perpetrators of much false logic. On the other hand, few of them are as sexist as is Donne's speaker.

Adding these two poems helped me improve the match between the students and the test. On the posttest 88 people fell above the second most difficult question. Because the instruction improved students' ability to interpret irony in poetry, without the addition of these more difficult poems, the match between the test and the population would be much worse instead of essentially the same.

The non-ironic items also contain a substantial variation of difficulty. The data support my theory that it is easier to understand non-ironic items in poems that contain no irony than it is to understand non-ironic items in poems that contain some irony. The mean level of difficulty for the items in poems that featured no irony was -1.041 while he mean of non-ironic items in poems that contain irony was .149. This difference is statistically significant ($t=2.76$, $p<.05$).

On balance, it appears that an understanding of irony begins by recognizing what is not ironic. As the ability to understand irony increases, readers are better able to reconstruct ironic meanings in increasingly

difficult poems and to reconstruct ironic meanings that challenge their own beliefs and behaviors.

Understanding the Instruction

Rasch analysis provides unique benefits in evaluating the instruction. One advantage is that logit measures give a clear sense of the magnitude of students' change. Table 3 contains summary statistics from each of the groups. The students who received the direct instruction improved on average .49 logits from the pretest to the posttest, which means the odds for success on the average ironic item for the average student increased from 1.2:1 to 2.0:1, an incease of 67 percent. The students who received the tacit instruction improved .44 logits, which means the odds for success on the average ironic item for the average student increased from .96:1 to 1.5:1, an incease of 56 percent. The students who received no treatment performed worse by .04 logits, which means the odds for success on the average ironic item for the average studen remained at 1.1:1.

Table 4 indicates the within groups results of the repeated measures analysis of variance. The analysis of variance indicates that the difference in the effect of the three treatments is statistically significant ($p < .001$). The analysis of variance further establishes that these differences are a function of the difference among treatments and not a function of the differences among the four teachers or among the interaction between

teachers and treatments.

Insert Table 4 about here.

To specify the source of the difference, I did post hoc tests on the the effect of the treatments on adjacent trials using Fisher F-tests. The performance of both the direct group and the tacit group was significantly superior to the performance of the no treatment group ($p < .001$). However, there was no statistically significant difference between the performance of the direct and tacit groups ($p = .761$).

In addition to giving a clear sense of how much students improved, Rasch analysis also allows one to take a closer look at the effects of a tretment. Because it calculates a standard error of measurement for each individual, Rasch analysis allows researchers to examine the performance of individual students. To be sure, we want to know how a curriculum is affecting groups of students. However, we never want to lose sight of how it is affecting the individuals within these groups. Therefore, I examined how the direct and tacit treatments affected students at each grade level. Table 5 displays the number of students receiving the direct and tacit treatments whose logit measure changed by more than one standard error from the pretest to the posttest.

Insert Table 5 about here.

Seven of the 26 ninth-grade students who received direct instruction improved by two or more standard errors. Seven additional students improved by one or more standard errors. Two students performed worse by two or more standard errors. Given the assumption that the posttest population is likely to have a higher mean than the pretest group, a change greater than one standard error indicates that there is approximately an 84 percent likelihood that the change is not a result of chance. While this is not a traditional level of statistical significance, it is a noteworthy change for an individual, particularly since the standard error of a short test is relatively high, especially at the top of the scale. A change of two or more standard errors indicates that there is approximately a 98 percent likelihood that the change is not a result of chance, a change almost always regarded as statistically significant. These changes occurred among students scoring the range of the test scores, indicating that the treatment did not have a differential effect depending upon the initial ability of the student.

In contrast, only 1 of the 22 ninth-grade students who received the tacit treatment improved by two or more standard errors. Five additional students experienced a change of more than one standard error between the pretest and the posttest. Once again, there did not appear to be a differential effect of the instruction based on the ability of the students.

In the tenth-grade direct classes, 4 of the 36 students improved by two or more standard errors, and 11 additional students improved by one or more standard errors. One student performed worse by more than one standard error. Five of the 39 tenth-grade students who received the tacit instruction improved by two or more standard errors, and 12 additional students improved by one or more standard errors.

In the eleventh-grade classes, 2 of the 24 students who received the direct instruction improved by two or more standard errors. Seven additional students improved by one or more standard errors. One of the 25 eleventh grade students who received the tacit instruction improved by two or more standard errors. Eight additional students improved by one or more standard errors. Both treatments had one student whose score declined by more than one standard error. As with the other grade levels, the effect of the treatment did not appear to depend on the ability of the student.

The ninth graders showed by far the largest differences between the two methods. This makes sense, for ninth-grade students are the least

experienced readers in the study. Consequently, they have had fewer chances to develop their own interpretive strategies. They are also the group for whom the task was the most difficult. The mean pretest score for the ninth-grade students in the direct and tacit groups was .29. The mean pretest score for the tenth grade students in the direct and tacit groups was .50, while the mean pretest score for the eleventh grade students in the direct and tacit groups was .85. The data suggest that the direct method may be more useful than the tacit method in helping students cope with more difficult tasks. If a task is easy, one can do it naturally. Reflection on the process is not necessary. However, when a task is problematic, it may be important to have conscious control of the strategies one may use to accomplish it.

The direct method also appears to carry additional risk. Four students who received the direct instruction scored more than one standard error worse on the posttest whereas this was only true for one student who received the tacit instruction. Perhaps because readers have a limited capacity for attention, drawing attention to interpretive strategies may take away from the attention some students pay to the critical details of a text.

Logit measures help one understand the magnitude of a treatment's effect, and a standard error of measurement for individuals allows one to

examine the effect of a treatment on individual students. Another advantage of Rasch analysis is that fit statistics can provide insight into the effectiveness of a treatment. As I explained earlier, 15 students misfit when I used dichotomous scoring. When I analyzed these 15 msifits, I found that 13 of them were on posttests. I also found that the highest residuals came on the two poems that were not ironic. This made sense. I hypothesized that some students who had just finished a unit on irony might be expected to read ironically even when the text does not call for such a reading. To test this hypothesis, I recalibrated the statistics after eliminating the questions on the two poems that were not ironic. Thirteen of the 15 misfits were eliminated by this procedure. This suggests that the instruction should devote more time on what Booth (1974) calls "knowing when to stop."

Rasch analysis provided me a number of important benefits. It helped me design a reliable instrument, modify that instrument as needed, and determine an effective scoring system. It helped me gain important insights into the variable I was measuring. And it helped me evaluate the effectiveness of my experimental treatments by giving me a clear sense of the magnitude of students' change and allowing me to look closely at the performance of individual students. Evaluating the quality of students'

responses may not be a popular practice, but Rasch analysis makes it a profitable one.

Sources Cited

Bleich, David. (1975). Readings and feelings: An introduction to subjective criticism. Urbana, IL: NCTE.

Booth, W. (1974). A rhetoric of irony. Chicago: University of Chicago Press.

Cooper, C. (1985). Evaluating the results of classroom literary study. In C. Cooper (Ed.), Researching response to literature and the teaching of literature: Points of departure (pp. 307-332). Norwood, NJ: Ablex Publishing Corporation.

Fish, S. (1980). Is there is a text in this class? Cambridge: Harvard University Press.

Holland, N. (1975). Five readers reading. New Haven: Yale University Press.

Marshall, J. (1989). Patterns of discourse in classsroom discussions of literature. University at Albany, State University of New York, Center for the Learning and Teaching of Literature.

Smith, M. (1989). Teaching the interpretation of irony in poetry. Research in the Teaching of English, 23, 254-272.

Wright, B., & Stone, M. (1979). Best test design. Chicago: MESA Press.

Table 1

A CAalibration of Item Difficulties

"Dulce et Decorum Est" 3	-1.82
"Sonnet 18" 1	-1.56
"Dulce et Decorum Est" 1	-1.53
"Sonnet 18" 4	-1.23
"Sonnet 18" 2	-1.02
"An Unknown Citizen" 2	-.93
"Dulce et Decorum Est" 2	-.93
"'next to of course god america i" 3 (1)	-.59
"'next to of course god america i" 4 (1)	-.55
"Sonnet 130" 4 (1)	-.51
"Sonnet 130" 3	-.48
"'next to of course god america i" 2 (1)	-.47
"Dead Boy" 1	-.43
"Boom!" 3	-.38
"Base Details" 3 (1)	-.38
"Dulce et Decorum Est" 4	-.24
"Base Details" 1 (1)	-.14
"Sonnet 18" 3	.00
"Boom!" 2 (1)	.01
"Base Details" 4	.01
"Dead Boy" 3 (1)	.23
"Base Details" 2	.25
"Communitie" 4 (1)	.31
"Sonnet 130 2 (1)	.35
"Boom!" 1 (1)	.62
"'next to of course god america i" 1 (1)	.63
"Communitie" 3 (1)	.75
"Communitie" 2 (1)	.76
"An Unknown Citizen" 3 (1)	.79
"Sonnet 130" 1 (1)	.84
"Dead Boy" 4 (1)	.89
"An Unknown Citizen" 1	.92
"An Unknown Citizen" 4 (1)	1.04
"Communitie" 1 (1)	1.15
"Dead Boy" 2 (1)	1.38
"Boom!" 4 (1)	2.26

Table 2

Summary Statistics of Item Difficulties

	Mean	s.d.
Ironic items	.45	.72
Non-ironic items	-.62	.73
Non-ironic items in ironic poems	-.15	.56
Non-ironic items in non-ironic poems	-1.04	.60

Table 3

Summary Statistics of Students' Logit Measures by Method and Grade Level

	Direct				Tacit				No			
Grade	9	10	11	total	9	10	11	total	9	10	11	total
n	26	36	24	86	22	39	25	86	24	34	22	80
Pretest												
M	.35	.74	.92	.67	.23	.29	.78	.42	.52	.49	.82	.59
SD	.52	.92	.76	.80	.56	.55	.76	.66	.63	.67	.78	.69
Posttest												
M	.99	1.20	1.28	1.16	.54	.80	1.23	.86	.56	.43	.72	.55
SD	.52	.70	.89	.82	.49	.70	.85	.74	.64	.48	.68	.59
Change	.64	.46	.36	.49	.31	.51	.45	.44	.04	-.06	-.10	-.04

Table 4

Results of Analysis of Variance

SOURCE	DF	MS	F	P
Teacher	3; 240	0.06	0.332	.802
Treatment	2; 240	3.34	17.949	.001
Teacher x Treatment	6; 240	0.17	0.901	.495
error		0.19		

Table 5

Changes Greater than One Standard Error in Logit Measures of Test Performance by Grade and Treatment

Grade	Treatment	n	Students changing more than one standard error		Students changing more than two standard errors	
			Increase	Decrease	Increase	Decrease
9	Direct	26	7	0	7	2
	Tacit	22	5	0	1	0
10	Direct	36	11	1	4	0
	Tacit	39	12	0	5	0
11	Direct	24	7	1	2	0
	Tacit	25	8	1	1	0

Appendix

The Test

Each of the following seven poems is followed by four statements. Please indicate whether you strongly agree (SA), agree (A), disagree (D), or strongly disagree (SD) with each statement. Please respond to every statement. Circle only one response to each statement.

BASE DETAILS

If I were fierce and bald and short of breath,
I'd live with scarlet Majors at the Base.
And speed glum heroes up the line to death.
You'd see me with my puffy, petulant face,
Guzzling and gulping in the best hotel,
Reading the Roll of Honor. "Poor young chap,"
I'd say--"I used to know his father well;
Yes, we've lost heavily in this last scrap."
And when the war is done and youth stone dead,
I'd toddle safely home and die--in bed.

Siegfried Sassoon

The author admires officers. SA A D SD

The author believes that enlisted men feel honored to fight in wars. SA A D SD

The author believes that officers have a genuine concern for their men. SA A D SD

The author believes that war is equally dangerous for officers and enlisted men. SA A D SD

THE UNKNOWN CITIZEN
(To JS/07/M/378
This Marble Monument
Is Erected by the State)

He was found by the bureau of statistics to be
One against whom there was no official complaint,
And all reports on his conduct agree
That, in the modern sense of an old-fashioned word, he was a saint,
For in everything he did he served the Greater Community.
Except for the War till the day he retired
He worked in a factory and never got fired,
But satisfied his employers, Fudge Motors Inc.
Yet he wasn't a scab or odd in his views,
For his Union reports that he paid his dues,
(Our report on his union shows it was sound)
And our Social Psychology workers found
That he was popular with his mates and liked a drink.
The press are convinced that he bought a paper every day
And that his reactions to advertisements were normal in every way.

Policies taken out in his name prove that he was fully insured,
And his Health-card shows he was once in a hospital but left it cured.
Both his producers research and High-Grade living declare
He was fully sensible to the advantages of the Installment Plan
And had everything necessary to the Modern Man,
A phonograph, a radio, a car and a frigidaire.
Our researchers into Public Opinion are content
That he had the proper opinions for the time of year;
When there was peace, he was for peace; when there was war, he went.
He was married and added five children to the population,
Which our eugenist says was the right number for a parent of his generation,
And our teachers report that he never interfered with their education.
Was he free? Was he happy? The question is absurd:
Had anything been wrong, we would certainly have heard.

W. H. Auden

The author believes that the citizen's job, views, and possessions make him a unique individual.
SA A D SD

The author believes that society and its institutions never complained about the citizen.
SA A D SD

The author believes the citizen was free. SA A D SD

The author believes that the reader should approve the kind of life the citizen led. SA A D SD

"next to of course god america i

"next to of course god america i
land of the pilgrims' and so forth oh
say can you see by the dawn's early my
country 'tis of centuries come and go
and are no more what of it we should worry
in every language even deefanddumb
thy sons acclaim your glorious name by gorry
by jingo by gee by gosh by gum
why talk of beauty what could be more beau-
tiful than these happy heroic dead
who rushed like lions to the roaring slaughter
they did not stop to think they died instead
then shall the voice of liberty be mute?"

He spoke. And drank rapidly a glass of water

e e cummings

The author admires the speaker's attitude towards America. SA A D SD

The author believes the speaker is a sincere religious person. SA A D SD

The author believes the happy heroic dead should have stopped to think about what they were doing. SA A D SD

The author believes that an understanding and appreciation of American history are important. SA A D SD

SONNET 130

My mistress' eyes are nothing like the sun;
Coral is far more red than her lips' red;
If snow be white, why then her breasts are dun;
If hairs be wires, black wires grow on her head.
I have seen roses damasked, red and white,
But no such roses see I in her cheeks;
And in some perfumes is there more delight
Than in the breath that from my mistress reeks.
I love to hear her speak, yet well I know
That music hath a far more pleasing sound;
I grant I never saw a goddess go;
My mistress, when she walks, treads on the ground.
And yet, by heaven, I think my love as rare
As any she belied with false compare.

William Shakespeare

The author criticizes the woman described in the poem because her voice is not as pleasing as music. SA A D SD

The author believes that the woman described in the poem is truly loved. SA A D SD

The author believes that poets should be realistic about their subjects. SA A D SD

The author believes that the woman described in the poem has unattractive eyes. SA A D SD

SONNET 18

Shall I compare thee to a summer's day?
Thou art more lovely and more temperate:
Rough wind do shake the darling buds of May,
And summer's lease hath all too short a date;
Some times too hot the eye of heaven shines,
And often is his gold complexion dimmed;
And every fair from fair sometimes declines,
By chance or nature's changing course untrimmed;
But thy eternal summer shall not fade,
Nor lose possession of the fair thou ow'st;
Nor shall death brag thou wander'st in his shade,

When in eternal lines to time thou grow'st:
So long as men can breathe, or eyes can see,
So long lives this, and this gives life to thee.

William Shakespeare

The author believes that the woman described in the poem is lovelier than summer.
SA A D SD

The author believes that poetry has the power to make love immortal. SA A D SD

The author believes that fair things in nature lose their fairness. SA A D SD

The author believes that the woman described in the poem is more temperate than summer.
SA A D SD

BOOM!
SEES BOOM IN RELIGION, TOO

Atlantic City, June 23, 1957 (AP).--President Eisenhower's pastor said tonight that Americans are living in a period of "unprecedented religious activity" caused partially by paid vacations, the eight-hour day and modern conveniences.

"These fruits of material progress," said the Rev. Edward L. R. Olsen of the National Presbyterian Church, Washington, have provided the leisure, the energy, and the means for a level of human and spiritual values never before reached."

Here at the Vespasian-Carlton, it's just one
religious activity after another; the sky
is constantly being crossed by cruciform
airplanes, in which nobody disbelieves
for a second, and the tide, the tide
of spiritual progress and prosperity
miraculously keeps rising, to a level
never before attained. The churches are full,
the beaches are full, and the filling-stations
are full, God's great ocean is full
of paid vacationers praying an eight-hour day
to the human and spiritual values, the fruits,
the leisure, the energy, and the means, Lord,
the means for the level, the unprecedented level,
and the modern conveniences, which also are full.
Never before, O Lord, have the prayers and praises
from belfry and phonebooth, from ballpark and barbecue
the sacrifices, so endlessly ascended.

It was not thus when Job[1] in Palestine
sat in the dust and cried, cried bitterly;
when Damien[2] kissed the lepers on their wounds
it was not thus; it was not thus
when Francis[3] worked a fourteen-hour day
strictly for the birds; when Dante[4] took
a week's vacation without pay and it rained

part of the time, O Lord, it was not thus.
But now the gears mesh and the tires burn
and the ice chatters in the shaker and the priest
in the pulpit, and Thy Name, O Lord,
is kept before the public, while the fruits
ripen and religion booms and the level rises
and every modern convenience runneth over,
that it may never be with us as it hath been
with Athens and Karnak[5] and Nagasaki[6],
nor Thy sun for one instant refrain from shining
on the rainbow Buick by the breezeway
or the Chris Craft with the uplift life raft;
that we may continue to be the just folks we are,
plain people with ordinary superliners and
disposable diaperliners, people of the stop'n'shop
'n'pray as you go, of hotel, motel, boatel,
the humble pilgrims of no deposit no return
and please adjust thy clothing, who will give to Thee
if Thee will keep us going, our annual
Miss Universe, for Thy Name's Sake, Amen.

Howard Nemerov

[1]the hero of an Old Testament book who keeps his faith despite the many terrible things that happened to him
[2]19th century missionary whpo cared for lepers. He eventually died of leprosy himself.
[3]a saint famous for his simple life and his way with animals
[4]a great Italian poet
[5]Egyptian town where the great ancient city of Thebes once stood
[6]the second city destroyed by an atomic bomb

The author believes that the spiritual progress of his society keeps rising. SA A D SD

The author admires the form that the religious devotions of his society take. SA A D SD

The author believes that the wealth of the modern world is greater than at any other time in history. SA A D SD

The author believes that if people are fortunate it is because God is watching out for them. SA A D SD

DULCE ET DECORUM EST

Bent double, like old beggars under sacks,
Knock-kneed, coughing like hags, we cursed through sludge,
Till on the haunting flares we turned our backs
And towards our distant rest began to trudge.
Men marched asleep. Many had lost their boots
But limped on, blood shod. All went lame; all blind;
Drunk with fatigue; deaf even to the hoots
Of tired, outstripped Five-Nines[1] that dropped behind.
Gas! Gas! Quick, boys!--An ecstasy of fumbling,
Fitting the clumsy helmets just in time;

But someone still was yelling and stumbling
And flound'ring like a man in fire or lime...
Dim, through the misty panes and thick green light,
As under a green sea, I saw him drowning.
In all my dreams, before my helpless sight,
He plunges at me, guttering, choking, drowning.

If in some smothering dreams you too could pace
Behind the wagon that we flung him in,
And watch the white eyes writhing in his face,
His hanging face, like a devil's sick of sin;
If you could hear at every jolt, the blood
Come gargling from the froth-corrupted lungs,
Obscene as cancer, bitter as the cud
Of vile incurable sores on innocent tongues,--
My friend, you would not tell with such high zest
To children ardent for some desperate glory,
The old Lie: Dulce et decorum est Pro patria mori.[2]

Wilfred Owen

[1]gas-shells
[2]It is sweet and fitting to die for one's country.

The author believes that being in a war is a horrible experience. SA A D SD

The author believes that some people see war as a means to achieve glory. SA A D SD

The author believes that the experience of war would change the attitude of those who glorify it. SA A D SD

The author believes that it is sweet and fitting to die for one's country. SA A D SD

DEAD BOY

The little cousin is dead, by foul subtraction,
A green bough from Virginia's aged tree,
And none of the country kin like the transaction,
Nor some of the world of outer dark, like me.

A boy not beautiful, nor good, nor clever,
A black cloud full of storms too hot for keeping,
A sword beneath his mother's heart--yet never
Woman bewept her babe as this is weeping.

A pig with a pasty face, so I had said,
Squealing for cookies, kinned by poor pretense
With a noble house. But the little man quite dead,
I see the forbears' antique lineaments.

The elder men have strode by the box of death
To the wide flag porch, and mutterring low send round

The bruit[1] of the day. O friendly waste of breath!
Their hearts are hurt with a deep dynastic wound.

He was pale and little, the foolish neighbors say;
The first-fruits, saith the Preacher, the Lord hath taken;
But this was the old tree's late branch wrenched away,
Grieving the sapless limbs, the shorn and shaken.

John Crowe Ransom

[1] news

The author believes that the boy was from an old, established family. SA A D SD

The author shares the feeling of mourning that the relatives demonstrate. SA A D SD

The author admires the boy's family. SA A D SD

The author finds consolation in the preacher's words. SA A D SD

COMMUNITIE

Good wee must love, and must hate ill,
For ill is ill, and good good still.
But there are things indifferent,
Which wee may neither hate, nor love,
But one, and then another prove,
As wee shall find our fancy bent.

If then at first wise Nature had
Made women either good or bad,
Then some wee might hate, and some chuse,
But since shee did them so create,
That wee may neither love, nor hate,
Onely this rests, All, all may use.

If they were good it would be seene,
Good is as visible as greene,
And to all eyes it selfe betrays:
If they were bad, they could not last,
Bad doth it selfe, and others wast,
So, they deserve nor blame, nor praise.

But they are ours as fruits are ours,
He but that tasts, he that devours,
And he that leaves all, doth as well:
Chang'd loves are but chang'd sorts of meat,
And when he hath the kernell eate,
Who doth not fling away the shell?

John Donne

The author admires the reasoning the speaker uses in advancing his beliefs. SA A D SD

The author believes that it is morally right treat things that are neither good nor bad in any way we choose. SA A D SD

The author believes that it is easy to detect goodness. SA A D SD

The author endorses the speaker's attitude towards women. SA A D SD